CLONING AND SEQUENCE ANALYSIS OF BOVINE β-CASEIN cDNA

Rafael Jimenez-Flores, Young C. Kang and Tom Richardson

Department of Food Science and Technology
University of California
Davis, CA 95616

SUMMARY: A bovine β-casein cDNA clone was isolated from a cDNA library prepared from mammary gland mRNA. Sequence analysis revealed 25 nucleotides (nt) of the 5' noncoding region, 672 nt of the complete sequence coding and a 3'region of approximately 500 nt. When the nucleotide sequence of bovine β-casein cDNA is compared to rat β-casein cDNA (5), a high degree of homology is observed in the first 100 nt corresponding to the signal peptide of the pre-β-caseins.

Bovine caseins represent a particularly well characterized group of proteins, and the amino acids of the four phosphoproteins comprising this system, α_{s1}-, α_{s2}-, β- and κ-caseins, have been sequenced (1-4).

Nucleotide sequence analysis has been performed on the caseins from rat (5,6), mouse (7), guinea pig (8,9) and sheep (10). Bovine α_{s1}- and κ-casein cDNAs have been completedly sequenced (11,12), and the bovine α_{s1}-casein structural gene has been isolated and compared to that of rat α-casein (19). However, only fragments of the cDNA coding for bovine β-casein have been characterized (13,14).

Presented in this paper is the complete nucleotide sequence of the coding region of a β-casein cDNA clone isolated from bovine mammary gland mRNA.

Cloning of β-casein cDNA has been performed with the intention of expressing the protein in yeast cells and, upon nt sequence manipulations, producing changes in the structure and functionality of the protein in food-related applications.

MATERIALS AND METHODS

Escherichia coli RRI and JM 105 strains were used for constuction of the cDNA library and sequencing of the cloned plasmids, respectively. Total

nucleic acids were isolated from bovine lactating mammary gland, and poly (A) RNA was purified by oligo (dt) cellulose chromatography (15). cDNAs were synthesized according to Gubler and Hoffman (16) and, after homopolymer tailing, were inserted into the *Pst I* site of pBR322. Following transformation, the cDNA library was screened by coloning hybridization (15) using a mixture of the 16 different 14-base oligonucleotides coding for all possible degeneracies in the 143-147 amino acid sequence of β-casein. Restriction enzyme maps and Southern blots were performed as described in (15). Selected clones were sequenced according to the method of Sanger et al. (17) and by the double stranded DNA method of Chen and Seeburg (18).

RESULTS AND DISCUSSION

A total of 87 clones were obtained from 6000 colonies in the library, using the radiolabelled oligonucleotide probes. Further restriction enzyme analysis of each clone with *Pst I* and *Bst NI* allowed selection of 20 clones based on the fragmentation patterns observed after agarose gel electrophoresis. Southern blots were obtained on the inserted fragments of each of the 20 clones, which were excised by digestion with *Pst I*. A single clone containing a 1.2 Kb insert was selected and its *Hae III* fragments were subcloned into M13 mp19 for sequencing. The sequencing strategy is shown in Fig. 1. It was convenient to subclone fragments of the orignal insert into

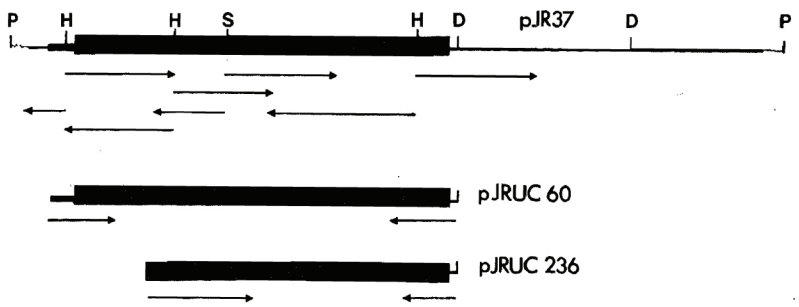

Figure 1. Strategy for sequencing and subcloning.

The top line represents the 1.2 Kb nt seqment of the bovine β-casein cDNA. Only the restriction sites used in subcloning are shown. The Pst I sites are generated by the cloning procedure. The signal peptide and the coding region are shown by the thicker lines. The arrows under pJR 37 indicate sequencing results from the method of Sanger et al. (17) and the arrows under pJR 60 and pJR 237 indicate sequencing results from the Chen and Seeburg method (18).
(Restriction enzymes P = *Pst I* S = *Sca I* H = *Hae III* D = *Dra I*)

```
                10                  20                  30                  40                  50                  60
                *                   *                   *                   *                   *                   *
A CTT GGA AAA AAG GAA TTG AGA GCC ATG AAG GTC CTC ATC CTT GCC TGC CTG GTG GCT CTC
                                  Met Lys Val Leu Ile Leu Ala Cys Leu Val Ala Leu

                70                  80                  90                  100                 110                 120
                *                   *                   *                   *                   *                   *
GCC CTT GCA AGA GAG CTG GAA GAA CTC AAT GTA CCG GGT GAG ATT GTG GAA AGC CTT TCA
Ala Leu Ala Arg Glu Leu Glu Glu Leu Asn Val Pro Gly Glu Ile Val Glu Ser Leu Ser

               130                 140                 150                 160                 170                 180
                *                   *                   *                   *                   *                   *
AGC AGT GAG GAA TCT ATT ACA CGC ATC AAT AAG AAA ATT GAG AAG TTT CAG AGT GAG GAA
Ser Ser Glu Glu Ser Ile Thr Arg Ile Asn Lys Lys Ile Glu Lys Phe Gln Ser Glu Glu

               190                 200                 210                 220                 230                 240
                *                   *                   *                   *                   *                   *
CAG CAG CAA ACA GAG GAT GAA CTC CAG GAT AAA ATC CAC CCC TTT GCC CAG ACA CAG TCT
Gln Gln Gln Thr Glu Asp Glu Leu Gln Asp Lys Ile His Pro Phe Ala Gln Thr Gln Ser

               250                 260                 270                 280                 290                 300
                *                   *                   *                   *                   *                   *
CTA GTC TAT CCC TTC CCT GGG CCC ATC CCT AAC AGC CTC CCA CAA AAC ATC CCT CCT CTT
Leu Val Tyr Pro Phe Pro Gly Pro Ile Pro Asn Ser Leu Pro Gln Asn Ile Pro Pro Leu

               310                 320                 330                 340                 350                 360
                *                   *                   *                   *                   *                   *
ACT CAA ACC CCT GTG GTG GTG CCG CCT TTC CTT CAG CCT GAA GTA CTG GGA GTC TCC AAA
Thr Gln Thr Pro Val Val Val Pro Pro Phe Leu Gln Pro Glu Val Leu Gly Val Ser Lys

               370                 380                 390                 400                 410                 420
                *                   *                   *                   *                   *                   *
GTG AAG GAG GCT ATG GCT CCT AAG CAC AAA GAA ATG CCC TTT CCT AAA TAT CCA GTT GAG
Val Lys Glu Ala Met Ala Pro Lys His Lys Glu Met Pro Phe Pro Lys Tyr Pro Val Glu

               430                 440                 450                 460                 470                 480
                *                   *                   *                   *                   *                   *
CCC TTT ACT GAA AGC CAG AGC TTG ACT TTG ACT GAT GTT GAA AAT CTT CAT TTG CCT CCT
Pro Phe Thr Glu Ser Gln Ser Leu Thr Leu Thr Asp Val Glu Asn Leu His Leu Pro Pro

               490                 500                 510                 520                 530                 540
                *                   *                   *                   *                   *                   *
CTC TTG CTC CAG TCT TGG ATG CAT CAG CCT CAC CAG CCT CTT CCT CCA ACT GTC ATG TTT
Leu Leu Leu Gln Ser Trp Met His Gln Pro His Gln Pro Leu Pro Pro Thr Val Met Phe

               550                 560                 570                 580                 590                 600
                *                   *                   *                   *                   *                   *
CCT CCT CAG TCC GTG CTG TCC CTT TCT CAG TCC AAA GTC CTG CCT GTT CCC CAG AAA GCA
Pro Pro Gln Ser Val Leu Ser Leu Ser Gln Ser Lys Val Leu Pro Val Pro Gln Lys Ala

               610                 620                 630                 640                 650                 660
                *                   *                   *                   *                   *                   *
GTG CCC TAT CCC CAG AGA GAT ATG CCC ATT CAG GCC TTT CTG CTG TAC CAG CAG CCT GTA
Val Pro Tyr Pro Gln Arg Asp Met Pro Ile Gln Ala Phe Leu Leu Tyr Gln Gln Pro Val

               670                 680                 690                 700                 710                 720
                *                   *                   *                   *                   *                   *
CTC GGT CCT GTC CGG GGA CCT TTC CCT ATT ATT GTC TAA GAG GAT TTT AAA GGA ATG CCC
Leu Gly Pro Val Arg Gly Pro Phe Pro Ile Ile Val ---

               730                 740                 750
                *                   *                   *
CTC TAT TTT GAA TTG ACT GCG ACT GAA ATA TGG C
```

Figure 2. The nt sequence of bovine β-casein and its translated amino acids shown in the 5' to the 3' direction. The underlined regions correspond to the signal peptide (16 to 70) and to the phosphorylation sites.

the expression plasmid pUC 18 in order to obtain sufficient nucleotide sequences adjacent to the homopolymer tails.

The fragments sublconed into the *SMA I* site of the cloning region of pUC18 were initially digested with the exonuclease *Bal 31* and then cut

Dra I (isosquitzouer of *Aka III*). This yielded a population of fragments of different sizes as depicted in Fig. 1 (pJR 60 and pJR 23). This facilitated completion of the cDNA sequence in both directions. The complete translated coding region, including the 15 aminoacids of the signal peptide are shown in Fig. 2.

Comparison of the coding region of Bovine β-casein using a forward homology matrix with the reported coding sequence of rat β-casein (5) showed a

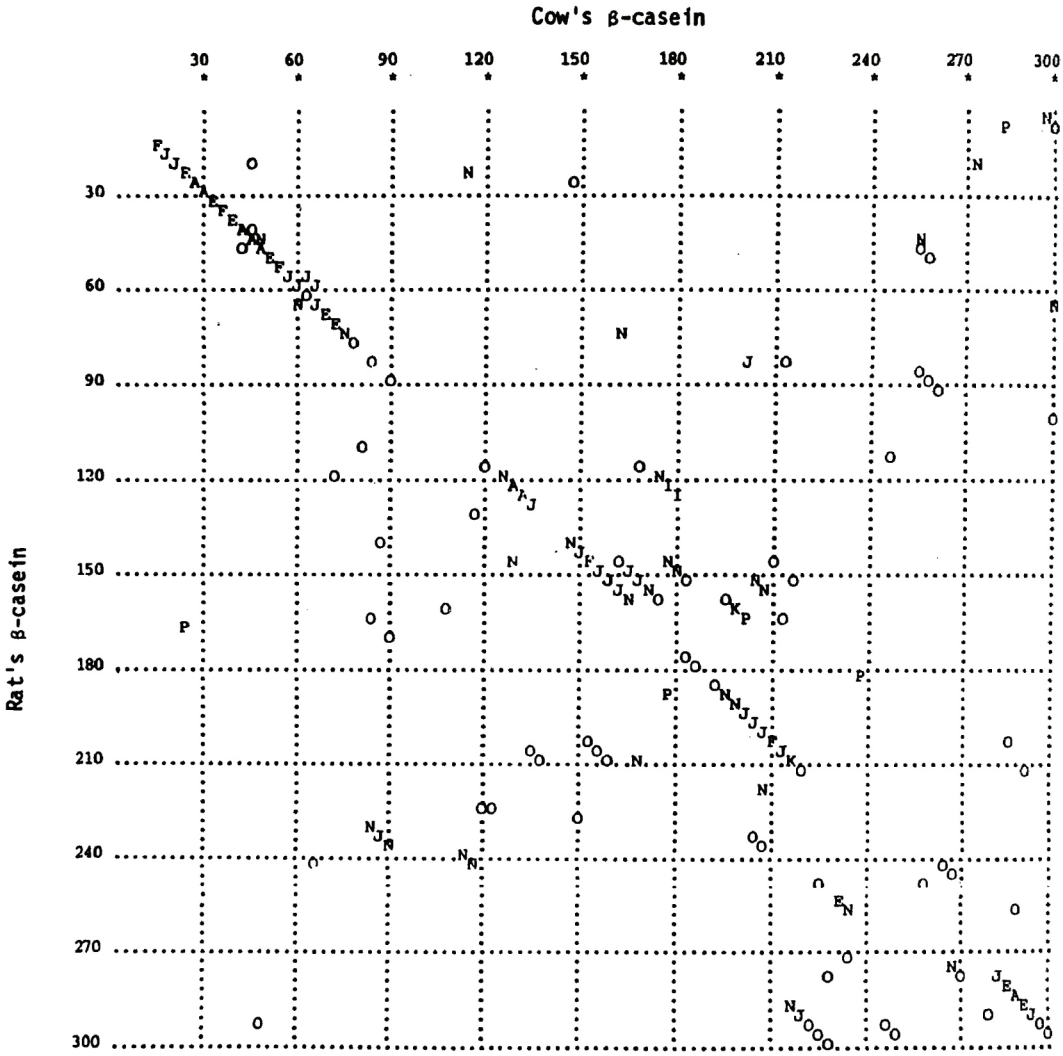

Figure 3. A computer generated forward homology matrix of the first 300 nt of bovine β-casein (x axis) and rat β-casein (y axis). Each point on the graph represents a base in the middle of a sequence of 10 nt of which 7 are homologous between the rat and bovine caseins. The plot for bases 20 to 75 represents the high homology of the signal peptides and the plot for bases 120 to 210 represents the homology at the phosphorylation sites.

very high homology of the regions corresponding to the signal peptide, and the phosphorylated segments of the caseins (Fig. 3). The nucleotides corresponding to these regions are underlined in Fig. 2. The Bal 31 generated fragments were screened for β-casein expression in E. coli using pUC 18 as the expression vector and IPTG induction. Preliminary results indicate the presence of β-casein as detected by western blotting (data not shown).

ACKNOWLEDGEMENTS

This research was supported by the Peter J. Shields Endowment Fund, the Dairy Research Foundation, the National Dairy Promotion and Research Board and by the College of Agricultural and Environmental Sciences, University of California, Davis, CA 95616.

REFERENCES

1. Mercier, J-C., Grosclaude, F., Ribadeau-Dumas, B. (1971) Eur. J. Biochem. **23**, 41-51.
2. Brignon, G., Ribadeau-Dumas, B., Mercier, J-C., Pelissier, J. P., Das, B. C. (1977) FEBS Lett. **76**, 274-279.
3. Ribadeau-Dumas, B., Brignon, G., Grosclaude, F., Mercier, J-C. (1972) Eur. J. Biochem. **25**, 505-514.
4. Mercier, J.-C., Bignon, G., Ribadeau-Dumas, B. (1973) Eur. J. Biochem. **35**, 222-235.
5. Blackburn, D. E., Hobbs, A. A., Rosen, J. M. (1982) Nucl. Acids Res. **10**, 2295-2307.
6. Hobbs, A. A., Rosen, J. M. (1982) Eur. J. Biochem. **125**, 131-141.
7. Hennighausen, L. G., Sippel, A. E. (1982) Eur. J. Biochem. **125**, 131-141.
8. Hall, L., Laird, J. E., Craig, R. K. (1984) Biochem. J. **222**, 561-570.
9. Hall, L., Laird, J. E., Pascall, J. C., Craig, R. K. (1984) Eur. J. Biochem. **138**, 585-589.
10. Mercier, J-C., Gaye, P., Saulier, S., Hue-Delabaie, D., Villotte, J-L. (1985) Biochemie. **67**, 959-971.
11. Stewart, A. F., Willis, I. M., MacKinlay, A. G. (1984) Nucl. Acids Res. **12**, 3895-3907.
12. Nagao, M., Maki, M., Sasaki, R., Clubia, H. (1984) Agric. Biol. Chem. **48**, 1663-1667.
13. Willis, I. M., Stewart, A. F., Caputo, A., Thompson, A. R., MacKinlay, A. G. (1982) DNA. **1**, 375-386.
14. Ivanov, V. N., Kersbulite, D. R., Bayer, A. A., Akhondova, A. A., Sulimova, G. E., Judinkova, E. S., Gorodestsky, S. I. (1984) Gene. **32**, 381-388.
15. Maniatis, T., Fritsch, E. F., Shambrook, J. (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. Gubler, V., Hoffman, B. J. (1983) Gene **25**, 263-269.
17. Sanger, F., Nicklen, S., Coulson, A. R. (1977) Proc. Nat. Acad. Sci. USA, **74**, 5463-5467.
18. Chen, E. Y., Seeburg, P. H. (1985) DNA **4**, 165-170.
19. Yu-Lee, L-Y., Richter-Maun, L., Couch, C. H., Stewart, A. F., Mackinlay, G., Rosen, J. M. (1986) Nucl. Acids Res. **14**, 1883-1902.
20. Mercier, J-C., Gaye, P. (1982) J. Dairy Sci. **65**, 299-316.